A NOTE ON A BIASED ESTIMATOR IN SAMPLING WITH PROBABILITY PROPORTIONAL TO SIZE WITH REPLACEMENT

By M. T. Subrahmanya

Indian Statistical Institute, Calcutta

A finite population of N units $U_1, U_2, \cdots, U_N$ is considered. Let Y_i denote the value of the characteristic under study for the ith unit. It is desired to estimate the total $Y = \sum_{i=1}^{N} Y_i$ on the basis of a sample.

When sampling with equal probability, with replacement after each draw, the total Y may be estimated unbiasedly by

$$\hat{Y}' = (N/n) \sum_{i=1}^{n} y_i,$$

where y_i is the value recorded at the ith draw. Basu (1958) has shown that the estimator $\hat{Y}_d'$ based on distinct units in the sample, namely,

$$\hat{Y}_d' = (N/n_d) \sum_{i=1}^{n_d} y_{(i)}$$

is unbiased for Y and has on the average a smaller mean-square error than $\hat{Y}'$. Here n_d stands for the number of distinct units in a sample of size n and the suffix (i) indicates the ith distinct unit in the sample.

When sampling with probability proportional to size, with replacement after each draw (pps for brevity), the total Y may be estimated unbiasedly by

$$\hat{Y} = n^{-1} \sum_{i=1}^{n} (y_i/p_i),$$

where p_i is the probability of selecting the unit occurring at the ith draw. Basu (1958) presents for this design an unbiased estimator Y_d superior to $\hat{Y}$, which makes use only of the values recorded for the n_d distinct units. This estimator $\hat{Y}_d$ is—contrary to what is stated in Dalenius (1962) and in some other papers—*not* identical with

$$\hat{y}_d = n_d^{-1} \sum_{i=1}^{n_d} [y_{(i)}/p_{(i)}].$$

We stress here that $\hat{y}_d$ is not the same as $\hat{Y}_d$, except in some special cases. For instance, when $n_d = n - 1$,

$$\hat{Y}_d = n^{-1} \sum_{i=1}^{n-1} y_{(i)} / \sum_{i=1}^{n-1} p_{(i)} + [(n-1)/n]\hat{y}_d.$$

In fact, $\hat{y}_d$ is in general not unbiased. In view of the relative simplicity with which $\hat{y}_d$ may be computed, we will study the properties of $\hat{y}_d$ in some detail.

To begin with, we make the following remarks:

(a) $\hat{Y}$, $\hat{Y}_d$ and $\hat{y}_d$ are identical when $n = 1$ or 2 or when $Y_i/P_i = Y$ for $i = 1, 2, \cdots, N$.

(b) $\hat{Y}_d$ and $\hat{y}_d$ are identical when $p_i = N^{-1}$ for $i = 1, 2, \cdots, N$.

Received 13 October 1965; revised 4 February 1966.